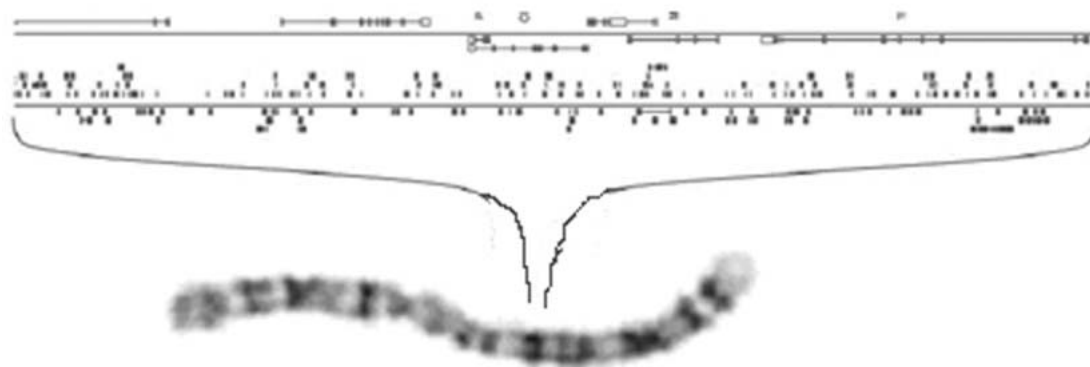


Från delar till helheter

- lärdomar från den moderna genomforskningen.

av Henrik Brändén



Förord

Då forskarvärlden för femton år sedan bestämde sig för att utröna hela DNA-sekvensen hos människan och en rad andra livsformer fanns åtminstone två skäl till att många biologer var entusiastiska:

För det första hoppades de att utifrån människans genom kunna konstruera en katalog över alla våra proteiner. Och att denna katalog skulle göra det möjligt för forskarna att anta ett mer holistiskt angreppssätt när de närmade sig olika biologiska problem. Så att de kunde studera alla proteiner inblandade i en process, inte bara enstaka proteiner.

För det andra hoppades de att studier av olika genom skulle visa vilka skillnader det egentligen är på molekylnivå både mellan olika individer och mellan olika arter. Och att detta skulle kunna kasta nytt ljus över frågor kring hur genetisk variation uppkommer, hur evolutionen går till och varför våra genom, och därmed vi själva, ser ut som de gör.

Några år efter det sekvensbestämningen av människans genom slutförts kan vi konstatera att bägge dessa förhoppningar tycks vara på väg att infrias.

Här skissas den helhetsbild av människans och andra arters genom och uppsättningar av proteiner som nu växer fram. Huvuddelen av det stoff som presenteras är kunskap som vunnits med de nya storskaliga tekniker som växt fram med det humana genomprojektet. Där så behövts för sammanhangets och fullständighetens skull har jag dock fogat in även en del fakta som varit kända sedan tidigare och en del kunskap som vunnits med andra metoder.

Mina källor i arbetet har varit vetenskapliga artiklar där genom presenterats och analyserats, liksom de öppna databaser som forskare skapat under arbetet. Förhoppningsvis har jag inte missförstått alltför mycket när jag läst och syntetiserat

Stockholm i augusti 2006
Henrik Brändén

Innehåll:

1. Människans proteiner
2. Människans RNA-molekyler
3. Människans genom
4. Utveckling av genom
5. Skillnader mellan människors genom
6. Skillnader mellan människans och schimpansens genom
7. Skillnader mellan människans och gnagares genom
8. Skillnaden mellan däggdjurs och andra arters genom

1. Människans proteiner

Tack vara sekvensbestämningen av människans genom har forskarna för första gången fått en katalog över alla de gener i människan som beskriver proteiner. Det visade sig att människan (liksom andra däggdjur) har mellan 22 000 och 23 000 olika proteinbeskrivande gener. Antalet olika proteiner är dock mångdubbelt större, kanske i storleksordningen 100 000. De flesta av däggdjurens gener består nämligen av ett antal moduler (exoner), och det maskineri som tillverkar proteiner kan vid olika tillfällen ignorera olika sådana moduler, och därigenom åstadkomma flera olika proteiner från samma gen. Ett fenomen som kallas alternativ splitsning.

Proteinernas funktion

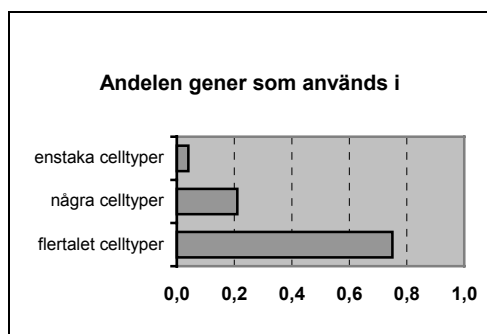
När forskarna väl lyckats etablera en katalog över alla gener som beskriver proteiner har de också kunnat börja studera och analysera hela denna uppsättning av proteiner för att lära sig förstå vilka funktioner de olika proteinerna har. Denna verksamhet kallas ofta för ”functional genomics”, eller funktionsgenomik. För var gen försöker forskarna då precisera svaret på fyra frågor:

- I vilka celler används genen?
- Var i cellen finns genens protein?
- I vilken biologisk process är proteinet inblandat?
- Vilken är proteinets biokemiska aktivitet?

Forskarna bygger därför upp databaser, där de för var gen lagrar den allt mer detaljerade kunskap som successivt växer fram. Utifrån dessa databaser kan man skissa en helhetsbild över människans proteinuppsättning – hennes proteom:

I vilka celler finns proteinerna?

Ungefär tre fjärdedelar av människans proteinbeskrivande gener används i alla eller de flesta av våra celler (figur 1.1). Bara en fjärdedel av dem används i ett mindre antal celltyper, och av



Figur 1.1. Andel gener som används i få respektive många celltyper.

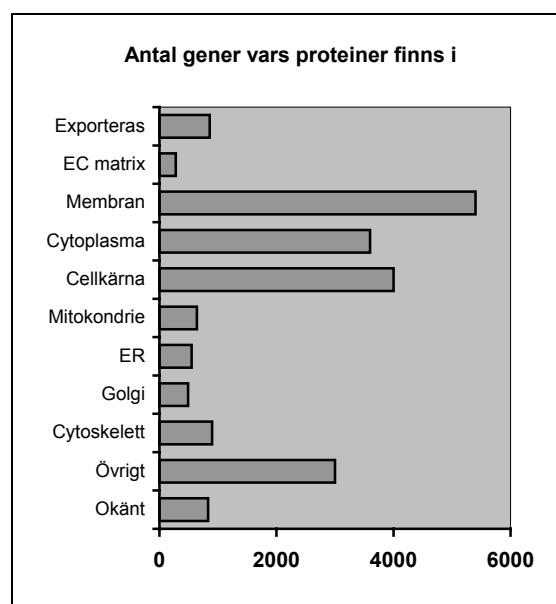
dem är det mindre än en tjugondel av våra gener som endast används i en enda celltyp. Detta har man kommit fram till med en teknik som kallas microarray (DNA-chips), som gör det möjligt att snabbt undersöka från vilka av människans alla gener en viss cell bildar budbärar-RNA.

Var i cellerna finns proteinerna?

Om man sedan frågar sig var i cellerna de olika proteinerna befinner sig ser man (figur 1.2) att runt 4000 av människans drygt 20 000 gener beskriver proteiner som befinner sig i cellernas cytoplasma. Proteiner från mer än 5000 olika gener går att hitta i cellens olika membran, och från 7000 gener i olika organeller. Av dessa står cellkärnan för överlägset flest. Över 1000 gener beskriver proteiner som transporteras ut ur cellerna, varav knappt 300 tycks bidra till att bygga olika former av extracellulära matrix, dvs strukturer utandför cellerna, som cellerna förankras i. Över 2000 gener vet vi idag beskriver proteiner, som ingår i stora komplex inne i cellerna bestående av tiotals eller hundratals olika proteiner. Mycket av denna kunskap har forskarna fått har forskarna fått genom att analysera sekvenser och leta beskrivningar för de så kallade signalpeptider i början av proteiners aminosyrekedja, som hjälper proteinet att hitta till en viss del av cellen.

Proteinernas biokemiska aktivitet.

En viktig aspekt av ett proteins funktion är dess biokemiska aktivitet (figur 1.3). Ungefär hälften av



Figur 1.2. Olika proteiners lokalisering i cellerna.



våra gener beskriver proteiner som har förmåga binda andra ämnen. En fjärdedel av dem beskriver proteiner som kan katalysera kemiska reaktioner, en femtondel beskriver proteiner som är transportörer. Ungefär samma andel beskriver proteiner som kan reglera avläsningen av olika gener och mindre än en hundradel sådana som kan generera rörelse. Sex beskriver chaperoner (proteiner som hjälper andra proteiner att veckla ihop sig korrekt) och tre beskriver proteiner som lagrar näring. En stor del av denna kunskap kommer från sekvensjämförelser, där man hittat nya gener som haft slående sekvenslikhet med tidigare kända enzymer, signalämnen etc. Vilket lett till hypoteser, som sedan varit lätta att testa i provrör.

Proteinernas biologiska roll

Den biokemiska aktiviteten är dock bara en sida av ett proteins funktion. En lika viktig fråga är vilken biologisk process proteinet är inblandat i. Huvuddelen av våra gener beskriver proteiner som deltar i någon av tre processer (figur 1.4):

- Ämnesomsättning
 - Fånga upp signaler, skicka dem vidare och integrera olika signaler till beslut om vad ett protein ska göra eller vilka proteiner som ska tillverkas.
 - Styra utvecklingen från ett nybefruktat ägg till färdigutvecklade kroppsceller lokaliserade på rätt ställe i kroppen.
- Därutöver är några hundra olika proteiner inblandade i att åstadkomma rörelse. Mindre än hundra proteiner behövs för blodkoagulering och endast fem för pigmentering.

Grunden för kunskapen om proteiners biologiska roll kommer från muterade djur där man sett att en

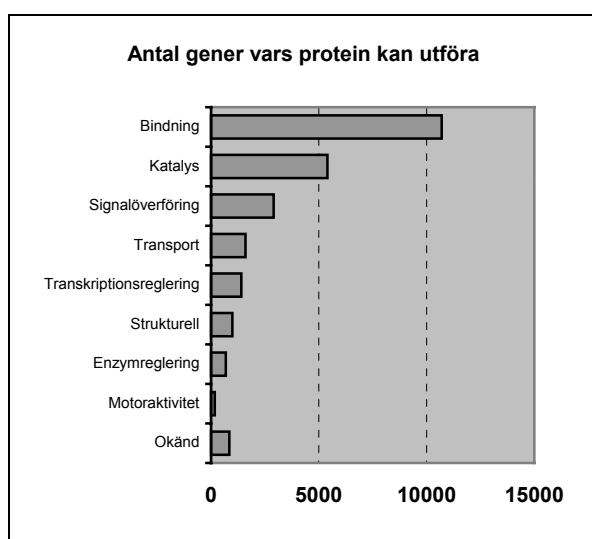
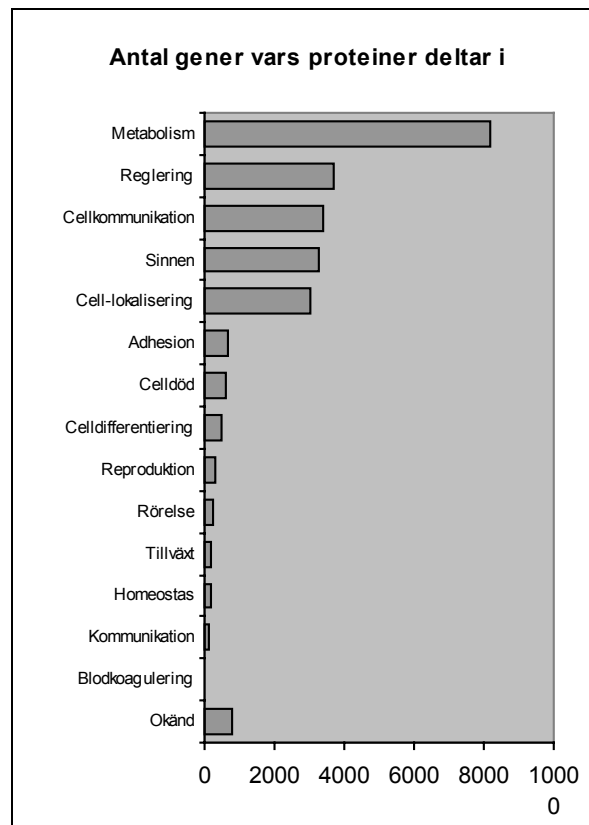


Fig 1.4 Proteiners biokemiska aktivitet.



Figur 1.4. Proteiners inblandning i olika biologiska processer.

viss funktion tagit skada om ett visst protein saknas. Om sedan andra experiment visar att fyra ytterligare proteiner binder till och alltid uppträder tillsammans med detta kan det vara reimligt att anta att även dessa är inblandade i samma process.

Att observera

När man betraktar de olika diagrammen ovan kan tre saker vara nyttiga att notera: (1) Diagrammen visar antal olika gener, inte antal proteinmolekyler eller sammanlagd proteinmassa. Exempelvis utgör en handfull olika muskelproteiner en betydande andel av vår proteinmassa. Medan bortåt tusen olika gener för luktreceptorer bidrar till mindre än en miljondel av antalet proteinmolekyler i kroppen. (2) Många av våra proteiner kan kombinera flera biokemiska aktiviteter, delta i flera biologiska processer eller befinna sig på flera olika ställen i cellen. Och kan därmed sätta sitt avtryck på flera ställen i samma diagram. (3) Dessa undersökningar är av tekniska skäl gjorda på gener, men många gener kan ge upphov till flera olika proteiner. Ibland vet vi att olika proteiner då bildas i olika celltyper eller hamnar på olika ställen, men vi vet inte hur vanligt detta är. Följaktligen vet vi inte hur det första diagrammet hade sett ut om det utgått från proteiner istället för gener.

2. Människans RNA-molekyler

Under de senaste åren har forskarnas syn på vår uppsättning RNA-molekyler ändrats dramatiskt. Speciellt vad gäller de fem procent av cellernas RNA som varken är ribosomalt- eller transfer-RNA.

Ribosomalt och transfer-RNA

Man har länge vetat att runt 80% av RNA:t i en människocell är ribosomalt RNA och att det i varje ribosom finns fyra olika ribosomala RNA-molekyler (som är 120, 160, 1900 och 4700 baser långa). Under det senaste årtiondet har forskarna insett att dessa RNA-molekyler inte som man tidigare trott bara har strukturella funktioner, utan också är intimt inblandade i ribosomens arbete. Vi vet idag att själva huvudsteget i translationen katalyseras av ribosomalt RNA, helt utan inblandning av protein.

Vidare visste man sedan tidigare att ungefär 15% av cellernas RNA utgörs av transfer-RNA. De finns i runt 40 olika grundutföranden, med olika tripletter på de ställen (antikodon) som binder till budbärar-RNA. Hos människan har man dock hittat så många som 500 olika gener för olika tRNA-molekyler. De flesta av dem är mellan 70 och 90 baser långa.

Proteinkodande RNA

Då proteinkodande gener avläses bildas i cellkärnan heterogent nukleärt RNA (hnRNA). Huvuddelen av den sekvens som då kopieras består dock av till synes meningslösa DNA-sekvenser, kallade introner. Dessa klipps bort från RNA-molekylen i en process som kallas splitsning. Denna sker medan avläsningen fortfarande pågår. Medan en RNA-molekyl fortfarande förlängs i sin ena ände kan sålunda intronerna i dess andra ände redan vara bortsplitsade. Även om längden på de gener som beskriver proteiner kan sträcka sig över

hundratusentals baspar är alltså de heterogena nukleära RNA-molekylerna bara en bråkdel så långa.

Bland annat genom splitsningen omvandlas det heterogena nukleära RNA:t till budbärar-RNA-molekyler som vandrar ut i cytoplasman där de möter ribosomer. En betydande del av budbärar-RNA-molekylerna består dock av regioner före och efter beskrivningen av proteiner, som kallas otranslaterade regioner (untranslated regions, UTR). Dessa har under senare år visat sig ha viktiga reglerande funktioner. Beroende på vilken sekvens de har kan olika proteiner binda till dem, som dels kan se till att budbärar-RNA:t hamnar i en speciell del av cellen, dels påverka hur snabbt ribosomen ska tillverka det protein, som budbäraren beskriver.

Innan människans genom sekvensbestämdes fanns bara fragmentarisk kunskap från enstaka gener om hur långa introner och otranslaterade regioner var. Idag vet vi (tabell 2.1) att en mänsklig gen i genomsnitt täcker 30 000 baspar på kromosomen, vilka alla avläses till RNA. Efter splitsning återstår i genomsnitt 2000 baser RNA, av vilka runt 1300 beskriver protein. De otranslaterade regionerna hos budbärar-RNA är således i genomsnitt 700 baser långa, och intronerna (som klipps bort vid splitsningen) är i storleksordningen 15 gånger längre än exonerna (de regioner av genen som behålls i budbärar-RNA:t). Varje gen har i genomsnitt 10 exoner. Variationerna kring dessa genomsnittliga siffror är dock mycket stora. Ett betydande antal gener sträcker sig över hundratusentals baspar samtidigt som flera procent av generna bara består av en enda exon.

Forskare har länge vetat att cellerna vid splitsningen kan ignorera en eller flera exoner, och att cellen har mekanismer som gör att den kan reglera vilka exoner som ska ignoreras vid vilka tillfällen. Detta ger cellerna möjlighet att göra flera olika proteiner/proteinvarianter från en gen, och kallas differentiell splitsning. Fram till år 2000 föreställde sig forskarna att detta var ett relativt ovanligt fenomen, som bara några enstaka gener utnyttjade. Undersökningar i genomforskningens kölvatten har dock visat att en stor andel av våra gener utnyttjar sådan differentiell splitsning.

En genomsnittlig mänsklig proteingen	
Omfattar på kromosomen	30 000 bp
Återstår efter splitsning (mRNA)	2 000 bp
Däruv beskrivande protein	1 300 bp
Vilket motsvarar	430 as
Antal exoner	10 st

Tabell 2.1. Genomsnittliga egenskaper hos mänskliga proteinbeskrivande gener.

Exakt hur ofta det förekommer är fortfarande oklart, men många forskare uppskattar att mer än hälften av proteingenerna kan splitsas på mer än ett sätt och att människans drygt 20 000 gener kan ge upphov till mellan 50- och 100 000 olika budbärar RNA-molekyler.

Andra långa RNA-molekyler

I cellerna finns dessutom ett mycket stort antal olika RNA-molekyler, som är långa (över 50 nukleotider) och har helt andra arbetsuppgifter än dem som hittills diskuterats. Några av dem har under de senaste årtiondena isolerats och studerats av forskarna. Bland annat har man karakteriserat RNA-molekyler som modifierar tRNA-molekyler, som deltar i splitsningen och som deltar i arbetet att leda vissa ribosomer till ERs membran (tabell 2.2).

Forskarna tycks dock hittills bara ha identifierat och karakteriserat en liten bråkdel av de långa icke proteinkodande RNA-molekyler som finns i våra celler. I genomforskningens kölvatten har man nämligen hos däggdjur funnit små mängder av tiotusentals olika slags långa RNA-molekyler, som inte beskriver proteiner. Dessa är ofta hundratals eller tusentals baser långa, de modifieras ofta och transporteras ut ur cellkärnan. I några fall vet man att de kan splitsas på olika sätt. När forskare undersökt ett litet urval av dessa RNA-molekyler såg de att flera av dem var nödvändiga för att en cell skulle överleva, och att även sådana som inte var absolut nödvändiga för att en cell skulle överleva ändå var inblandade i olika viktiga processer. Uppenbarligen finns alltså ett stort antal

Namn	Funktion
Small nucleolar RNA (snRNA)	Ingår i snRNP, som utför splitsning
Small nucleolar RNA (snoRNA)	Processar och modifierar baserna hos tRNA
7 SL RNA	Ingår i SRP, som hjälper ribosomer hitta till ER
Telomeras RNA	Ingår i telomeras, som förlänger telomerer
Rnase P	Trimmar tRNA

Tabell 2.2. Några kända långa RNA-molekyler

olika längre RNA-molekyler med viktiga arbetsuppgifter i cellerna. Att kartlägga dessa kan bli ett ännu mödosammare arbete för forskarna än kartläggningen av alla proteiners arbetsuppgifter.

Mikro-RNA

Under de senaste åren har forskarna också funnit att det i våra celler finns något som kallas mikro-RNA (miRNA). Dessa molekyler är dubbelsträngade, ca 20 baspar långa och spelar viktiga roller för att reglera produktionen av olika proteiner. Runt om i vårt genom finns hundartals, kanske tusen, gener för sådana mikro-RNA molekyler. Dessa gener är 70-90 baspar långa, och de RNA-molekyler som bildas från dem har regioner som kan baspara till varandra. Därigenom åstadkoms stam-öglestrukturer, och speciella enzymer finns i cellerna som ur dessa stam-öglestrukturer klipper ut en ungefär 20 baspar lång bit dubbelsträngat RNA. Som alltså kallas micro-RNA. Detta binder sedan till proteiner varvid ett komplex bildas, som påverkar produktionen av de proteiner, vars budbärar-RNA har samma eller liknande sekvens som mikro-RNAt.

Om någon del av en budbärar RNA-molekyl har exakt samma sekvens som micro-RNAt kan nämligen proteinerna vid mikro-RNAt klippa sönder budbärar RNA-molekylen. Och därmed effektivt förhindra att protein bildas. Om ett budbärar-RNA har en sekvens som bara är ungefär likadan som mikro-RNAts kan istället komplexet av mikro-RNA och proteiner binda till budbärar RNA-molekylen och hindra denna från att binda på rätt sätt till ribosomen. Vilket också hindrar cellen från att bilda proteinet.

Eftersom en mikro RNA-molekyl är mycket kort kan dess sekvens återfinnas i många olika gener. Genom att börja eller sluta tillverka ett enda slags mikro-RNA kan en cell i ett slag slå av eller på produktionen av ett stort antal olika proteiner. Forskare uppskattar att mikro-RNA kan vara med att reglera en betydande andel, kanske bortåt en tredjedel, av människans gener.

3. Människans genom

I enkel kromosomuppsättning består människans genom av 3,08 miljarder baspar. Av dessa utgör ungefär 200 miljoner baspar centromerer och telomerer (centralpunkter och ändregioner i kromosomerna), där DNA-molekylerna är mycket hårt packade kring histonproteinerna till så kallat heterokromatin. Dessa regioner tycks nästan uteslutande bestå av repetitiva sekvenser och man tror inte att det finns några gener där. Av tekniska skäl är dessa regioner nästan omöjliga att sekvensbestämma. Det som sägs i resten av detta avsnitt gäller därför de 2,88 miljarder baspar där DNA-molekylerna är löst packade, så kallat eukromatin

GC-innehåll, gentäthet och färgning

Av anledningar som forskarna inte förstår varierar proportionerna mellan GC- och AT-baspar kraftigt mellan olika arter. Tack vare sekvenseringen av människans genom har man kunnat bestämma andelen GC i vårt genom till 41%. Denna andel varierar dock betydligt även inom människans genom. I långa regioner kan andelen vara över 60%, i andra under 30%.

Sekvensbestämningen av vårt genom har visat att i de regioner där GC-halten är hög ligger generna tätt, och intronerna är relativt korta. Dessa regioner har en tendens att bli ljusa vid Giemsa-färgning (figur 3.1) av kromosomer. De band på kromosomerna som färgas mörkt av Giemsa är däremot vanligen AT-rika och genfattiga, och de gener som finns där har långa introner. Det är oklart ifall detta samband har någon djupare innebörd

Proteinkodande gener

De gener som beskriver proteiner ligger ojämnt utspridda på kromosomerna. Man kan tydligt urskilja kluster där proteinkodande gener ligger tätt, och andra områden där de ligger glest. Betraktar vi en bild av hela kromosom 16 (figur 3.2) ser man tydligt att de gentäta klustren tenderar att ligga på eller kring de ljusa banden, medan huvuddelen av de mörka kromosombanden är genfattiga. I samma bild visas en förstoring av en typisk gentät del av kromosomen, där man ser att generna ligger med korta mellanrum, näst intill invid varandra. Ibland överlappar de till och med. I de förstörade delarna av bilden kan man också se att huvuddelen av generna utgörs av introner och

att det finns en kraftig variation både i genernas längd, antalet exoner och intronernas längd.

Transposoner

En stor del (44%) av människans genom består av transposoner, DNA-sekvenser som har förmåga att hoppa eller skapa en kopia av sig själv, som den stoppar in på ett annat ställe i genomet. De allra flesta av dessa sekvenser är inte "levande" transposoner utan döda relikter från misslyckade hopp, där bara en del av transposonen lyckats infogas i genomet. I figur 3.2 ser man hur olika sådana (rester av) transposoner ligger relativt jämnt utspritt över hela kromosomen. Det finns dock både regioner som är ovanligt rika och som är ovanligt fattiga på transposoner. Man kan hitta områden om hundratusentals baspar där över 90% av DNA-sekvensen består av transposoner, och andra lika stora områden där de bara utgör enstaka procent. I den rätt typiska region som förstörats i figur 3.1 finns en stor mängd transposoner, som befinner sig såväl mellan generna som inne i deras introner. Transposoner finns hos människor och andra däggdjur av fyra olika klasser, som beskrivs närmare i en fördjupningsruta i slutet av detta avsnitt.

Duplikationer och pseudogener

I människans genom hittar man ett stort antal regioner där man tydligt ser att genomet relativt nyligen under evolutionen duplicerats, så att man fått två regioner som ursprungligen hade exakt samma sekvens, men sedan utvecklats åt olika håll. Inte mindre än 5% av genomet består av regioner som är flera tusen baspar långa och har duplicerats så nyligen (uppskattningsvis under de senaste 40 miljonerna år) att de har mer än 90% sekvens-

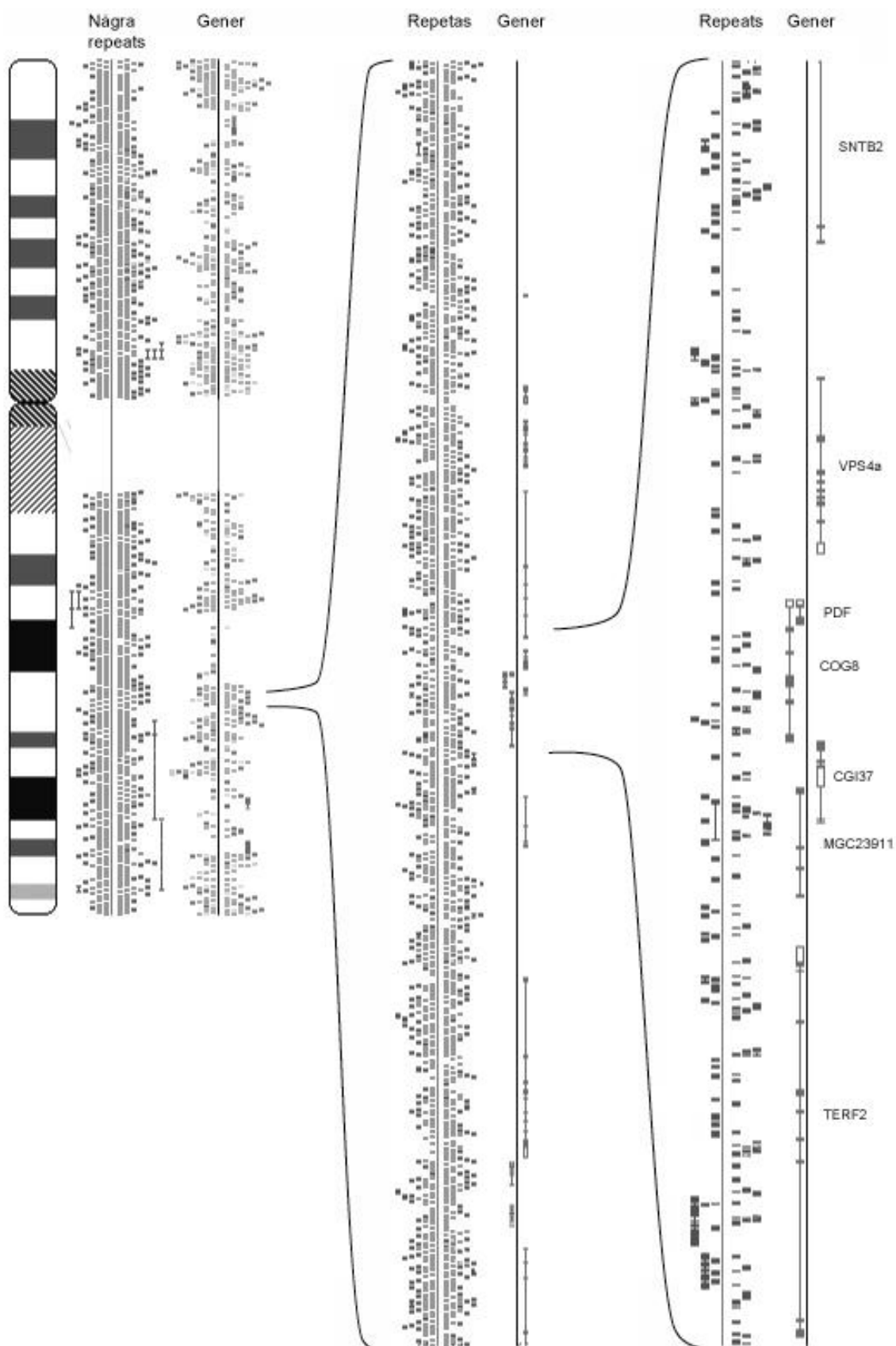


Figur 3.1. Giemsa-färgning av människans tre största kromosomar, nr 1-3. Foto David McDonald, Laboratory of Pathology of Seattle.

Hela kromosom 16 (89 miljoner baspar)

En miljon baspar

100 000 baspar



Figur 3.2 (föregående sida): Hela människans kromosom 16 och del av denna i olika upplösning. Vid repetitiva sekvenser markeras transposoner som tjocka streck (vilka i de lägre upplösningarna ofta inte går att urskilja från varandra) och dupliceringar som tunnare streck. Vid gener anges de exoner som återfinns i mRNA som boxar (kodande sekvens fylld box, ickekodande sekvens ofylld), introner som ett tunt streck. Bild: Henrik Brändén, Vetenskapsrådet.

likhet. I några fall ligger de duplicerade sekvenserna invid varandra men i de flesta fall ligger de på olika ställen. I nästan hälften av fallen ligger de duplicerade regionerna på olika kromosomer. I figur 3.2 kan man se flera duplikationer i de nedre regionerna av den del av bilden som visar hela kromosomen.

De regioner som kopierats innehåller en rad olika gener, och när en gen på detta sätt kopieras kan två saker hända med den: Endera utvecklas de två generna åt varsitt håll, så att exempelvis en gen för en luktreceptor ger upphov till två olika gener för två olika receptorer som känner igen två olika ämnen. Eller till att den ena genen råkar ut för en mutation som förstör genen. Exempelvis genom att det uppstår ett stoppkodon mitt inne i genen.. Sådana förstörda gener kallas pseudogener. De duplicerade delarna av våra kromosomer innehåller såväl många sådana pseudogener, som exempel på att duplicerade gener utvecklats åt olika håll.

Enkla repetitiva sekvenser

Utspritt i vårt genom finns också ställen där en kort sekvens (mindre än 500 baser) upprepas gång på gång efter varandra. Om den repetitiva enheten är så liten som 1-13 baspar (vilket oftast är fallet) talar man om stället som en mikrosatellit. Ungefär 3% av människans genom utgörs av sådana sekvenser. De ligger utspridda med ungefär 2000 basers mellanrum. Oftast befinner de sig mellan gener och i introner, men i enstaka fall kan de även finnas inne i de delar av gener som beskriver proteiner.

Funktionella delar av genomet.

Genom att jämföra människans genom med andra däggdjurs har forskarna kunnat identifiera de områden, som varit utsatta för positiv selektion (dvs selektion för att bevaras) under utvecklingen från en gemensam förfader. Detta har gjorts genom att börja med att beräkna vad som antas vara en neutral substitutionshastighet, som man mäter genom att studera frekvensen mutationer som inte antas vara utsatta för något selektivt tryck. (Endera mutationer inne i trasiga transposoner, eller mutationer som förvandlar en tripplett som beskriver en aminosyra till en annan tripplett som beskriver samma aminosyra.) Därefter har man helt enkelt undersökt vilka delar av genomen där betydligt färre mutationer än så bevarats. Forskarna har då funnit att 5-6% av genomet varit

utsatt för selektion att bevaras sedan människan och musen skildes från varandra för ungefär 70 miljoner år sedan.

Av dem utgör ungefär 1,5% beskrivningar av proteiner. Runt 1% beskriver otranslaterade delar av budbärar-RNA och idag kända icke-proteinkodande RNA-molekyler. Tre procent återstår, och forskarna antar idag att de består av (a) sådana reglerande sekvenser som påverkar när och hur mycket gener ska avläsas, (b) sekvenser som beskriver idag okända icke-kodande RNA-molekyler och eventuellt (c) sekvenser som är av betydelse för kromosomstrukturen.

3.8 Gåtan skräp-DNA

Betydligt gåtfullare än de enstaka procenten konserverad sekvens vars funktion vi ännu inte känner är dock det faktum att bortåt 95% av vårt genom utgörs av sekvenser som inte tycks vara utsatta för något selektivt tryck att bevaras, och som ibland kallas för skräp-DNA. Dessa sekvenser för med sig enorma kostnader för cellerna och därmed också för oss: När vi kopierar och underhåller våra gener går det åt 20 gånger mer energi och byggmaterial än det skulle ha gjort om vi saknade skräp-DNA. När gener avläses spjälkas 15 gånger fler energirika fostafrupper från ATP, GTP etc än vad som skulle skett om våra gener inte haft alla dessa långa introner.

Det finns några faktorer som möjligen skulle kunna förklara en liten del av detta till synes slöaktiga beteende:

- Introner kan behövas för att göra det möjligt att skapa ett stort antal olika proteiner från en och samma gen. Men inget talar för att sådan differentiell splitsning skulle kräva att intronerna är 15 gånger längre än exonerna. Många organismer klarar sig med introner som har ungefär samma storlek som exonerna.
- Det är rimligt att anta att en del av det icke-konserverade DNA:t kan ha någon typ av funktion som inte är beroende av sekvensen, till exempel att fungera som spacers, så att olika sekvenser hamnar på rätt ställe i förhållande till varandra. Men det förefaller svårt att tro att detta skulle kunna gälla för mer än några procent av genomet.
- Bortåt hälften av skräp-DNA:t utgörs av transposoner och är alltså parasitiskt till sin

karaktär. Man kan därför hävda att dessa sekvenser finns där för sin egens skull, inte för vår. Jämförelser mellan olika genom har dock visat att utrensningen av gammalt skräp ur genomet gått med olika hastighet hos olika utvecklingslinjer. Frågan reser sig därför varför vi inte gjort oss av med dessa parasitiska sekvenser snabbare än vi har.

Vår genom består sålunda till största delen av sekvenser som det inte funnits något selektivt tryck att bevara och för vilka man inte kan föreställa sig någon funktion som skulle kunna bidra till ökad reproduktiv framgång för individen. Detta DNA för med sig oerhörda kostnader för celler och individer. Det skulle ha varit möjligt att rensa bort dem i betydligt snabbare takt under utvecklingen mot människa. Men detta har inte skett.

En tanke som ofta förs fram är att detta skräp-DNA fungerar som experimentverkstad. Där finns sekvenser där nya mutationer och kombinationer av mutationer kan ske utan att några viktiga befintliga gener skulle slås ut. Som vi senare kommer att se har under evolutionen helt nya gener ofta skapats genom att bitar av gamla gener kopplats ihop med varandra på nya sätt. Ju större

avstånd mellan generna och ju större introner inne i generna, desto större möjligheter för sådana omkopplingar att ske. Har dessutom transposoner placerat nästan identiska kopior av sig själva i dessa mellanrum och introner ökar sannolikheten ytterligare för att sådana omkopplingar ska ske.

Problemet med en sådan förklaring till skräp-DNA är dock att sådana mekanismer för att skapa ökad genetisk variation bara leder till fördelar för hela populationer och detta på mycket lång sikt. Den idag dominerande tolkningen av evolutionsläran menar dock att det naturliga urvalet bara verkar på individnivå och bara kan selektera för egenskaper som leder till ökad reproduktiv framgång redan på någon eller några generationers sikt. Flertalet av de molekylärbioologer som reflekterat över denna paradox tror därför att fortsatta undersökningar av vårt genom med tiden kommer att avslöja idag okända funktioner även hos det som vi idag kallar skräp-DNA. Men några forskare leker istället med tanken att selektionen på individnivå för egenskaper som ger kortsiktiga överlevnadsfördelar för individen kan existera parallellt med en selektion på gruppnivå för egenskaper som bidrar till att skapa och upprätthålla genetisk variation.

Fördjupning: Transposoner

I människan och andra eukaryoter finns fyra klasser av transposoner:

DNA-transposoner: har i sina ändrar igenkänningssekvenser om 20-40 baspar, som kallas Inverted Repeats (IR). Mellan dem finns genen för ett enzym, kallat transposas, som kan binda till IR-sekvenserna, klippa ut transposonen och klistra in den på ett nytt ställe på en DNA-molekyl i cellen. En DNA-transposon är vanligen mellan 2 och 3 000 baspar lång. DNA-transposoner utgör 3% av människans genom. Huvuddelen av dessa sekvenser har dock muterats så att transposonen inte fungerar.

Retroviruslika transposoner: kallas också endogena retrovirus eller LTR-element. De har i ändarna igenkänningssekvenser som kallas "Long Terminal Repeat" (LTR). Mellan dem finns bland annat genen för enzymet omvänt transkriptas och ett integras-enzym. Om transposonen avläses bildas en RNA-molekyl som dessa proteiner kan binda till. De gör då en DNA-kopia av RNA-molekylen och klistrar in den på något ställe i genomet. Därmed har en kopia av transposonen lagts till cellens genom. Till sitt sätt att fungera påminner dessa mycket om retrovirus, och man kan beskriva dem som retrovirus utan höljeprotein. De är mellan 6 och 12 000 baspar långa och utgör 8% av människans DNA. Det mesta av även dessa sekvenser har dock sedan länge fått mutationer så att transposonen inte längre fungerar.

LINEs: (long interspersed elements) består av en styresekvens (promotor) som kan starta bildning av RNA, och fortsätter med gener för enzymet omvänt transkriptas och ett enzym som kan åstadkomma enkelsträngsbrott på en DNA-molekyl. Då avläsning startar från promotorn bildas RNA, som proteinerna kan gripa tag i. De slår sig ner någonstans på genomet, klipper av ena strängen av DNA-molekylen och tar tag i den andra änden och börjar förlänga den med RNA-molekylen som mall. När cellens maskineri för att sköta om DNA-molekylen sedan tar vid blir resultatet någon gång att hela transposonen fått en ny kopia av sig själv. Oftast avbryts dock DNA-syntesen innan hela RNA-molekylen är kopierad, varvid resultatet blir att en stympad transposon infogats. Som inte har förmåga att kopiera sig själv vidare. En LINE med förmåga att hoppa är 6 – 8000 baspar lång. Vårt genom består till 13% av LINEs, men de allra flesta av dem är korta, livlösa resultat av misslyckade försök till kopieringar.

SINEs (short interspersed elements) består av en promotor följt av en sekvens, som ger upphov till en RNA-molekyl, som någon LINE-transposons proteiner kan binda till. Så att dessa SINEs kan kopieras "på köpet" när en LINE-transposon är verksam. SINEs är typiskt mellan 100 och 400 baspar långa. De utgör 20% av människans genom.

4. Utveckling av genom

När forskare nu kunnat jämföra hela genom från olika varelser de lärt sig mycket om de genetiska förändringar som åstadkommer biologisk variation och därmed evolution.

Storskaliga kromosomförändringar

Ibland förekommer i cellerna storskaliga förändringar av arrangemanget av kromosomer: En kromosom kan dela sig i två, två kromosomer kan slås ihop eller en stor del av en kromosom kan flyttas till en annan. Vi har dålig förståelse av de molekylära mekanismer som ligger bakom sådana omflyttningar, men vi vet att de sker tillräckligt ofta i människoceller för att kunna orsaka en del medfödda sjukdomssyndrom och vissa cancerfall. Men sådana förändringar accepteras uppenbarligen sällan av det naturliga urvalet: Mellan människa och schimpans finns bara enstaka sådana storskaliga skillnader, och skillnaden mellan människans och musens kromosomuppsättningar kan förklaras av sammanlagt ca 200 storskaliga omstuvningar i utvecklingslinjerna från vår gemensamma förfader till dagens två arter.

Dupliceringar

Vi känner däremot en rad olika molekylära mekanismer genom vilka kortare segment av kromosomerna (mellan tusen och en miljon baspar) kan dupliceras, så att en del av genomet kommer att finnas i dubbel uppsättning. Sådana dupliceringar förekommer ofta: Jämför man två mänskliga kromosomuppsättningar hittar man tiotals regioner längre än 10 000 baspar som är duplicerade i den ena men inte den andra kromosomuppsättningen. Runt fem procent av människans genom består av duplikationer som skett de senaste 40 miljonerna år.

Sådana dupliceringar är oerhört viktiga för att skapa den genetiska variation det naturliga urvalet arbetar med. Om en gen dupliceras kan ju den ena kopian av genen förändras och utveckla nya funktioner, samtidigt som den andra kopian kan behålla sin ursprungliga funktion. Så att den levande varelsen kan skaffa sig förmågan att göra något nytt, utan att samtidigt förlora en tidigare förmåga (figur 4.1). För att bara ta ett exempel visar jämförelser av genomerna hos olika djur att utvecklingen av luktsinnet till stor del skett (och sker) genom dupliceringar av och mutationer i gener som beskriver olika luktreceptorer. Så att djur kan skaffa sig förmågan att upptäcka nya lukter utan att samtidigt förlora förmågan att urskilja andra. Något som är av stor betydelse

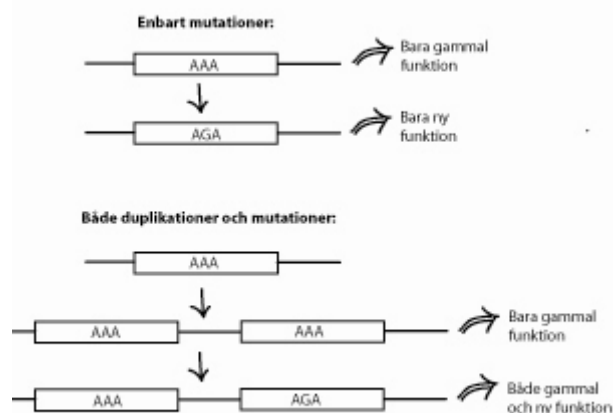
under evolutionen för att anpassa djur till nya eller förändrade miljöer, där det kan innebära en betydande överlevnadsfördel att kunna uppfatta lukten av exempelvis ett rovdjur, ett visst gift eller en viss föda. Duplikationer spelar således en viktig roll för att olika organismer snabbt ska kunna anpassa sig till förändrade miljöbetingelser.

Omstuvningar av domäner

Vi vet också att det finns en rad mekanismer med vilka kortare segment av genomet (några hundra eller några tusen baspar) kan kopieras eller flyttas omkring inom ett genom. Ibland kan detta leda till att beskrivningar av två olika mindre proteiner kan föras ihop och därmed slås ihop till en beskrivning av ett större protein, som i en aminosyrekedja kombinerar bägge de biokemiska aktiviteterna, som fanns i de ursprungliga proteinerna. Eller till att delar av olika gener förs ihop.

Under evolutionen har det åtskilliga gånger förekommit att beskrivningar av olika proteindomäner på detta sätt först ihop, så att det bildats gener för proteiner med komplexa aktiviteter. Exempelvis antas genen för steroidhormoner ha uppkommit genom att tre separata gener förts ihop som beskrev tre olika proteiner: ett som kan binda steroider, ett som kan binda DNA och ett som kan interagera med andra proteiner och därmed påverka avläsningen av en gen invid (figur 4.2).

Vi vet att mekanismer som åstadkommer sådana omstuvningar ständigt är aktiva i våra genom. Sannolikheten är dock liten för att en omstuvning



Figur 4.1. Duplikationer spelar en avgörande roll för att proteiner ska kunna utvecklas till att anta nya funktioner. Bild: Henrik Brändén, Vetenskapsrådet.

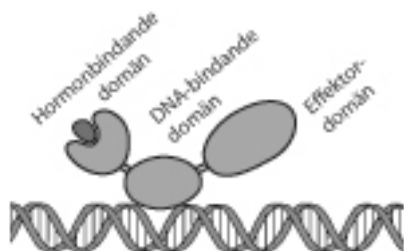
ska leda till uppkomsten av ett nytt protein som har en funktion som innebär en fördel för just den individ där omstuvningen skett. Uppkomsten av nya proteiner med nya kombinationer av domäner är därför en ovanlig evolutionär händelse. När man jämför olika ryggradsdjur med varandra hittar man bara enstaka exempel på sådana händelser. Däremot har sådana händelser varit viktiga under den utveckling som ledde fram till de olika djurklasserna. Vi ryggradsdjur har exempelvis dubbelt så många sätt att kombinera olika domäner med varandra än rundmaskar.

Nyskapande av domäner

Samma molekylära mekanismer som kan flytta beskrivningar av hela domäner kan naturligtvis också flytta enstaka exoner, så att det uppstår helt nya beskrivningar av domäner, med en ny inre tredimensionell arkitektur (dvs ett nytt arrangemang av alfahelixar och betasträngar inom domänen). Rimligen måste sådana omstuvningar ske ännu oftare i genomen än omstuvningar av hela domäner, men sannolikheten är ytterst liten att en sådan flytt av en eller ett par exoner skulle åstadkomma en ny domän som både är stabil och kan utföra något som innebär en fördel för sin bärare. Det händer därför oerhört sällan att nya fungerande domäner uppkommer och bevaras av det naturliga urvalet. Hos däggdjur finns exempelvis bara några tiotal typer av domäner som inte har hittats hos ryggradslösa djur.

Insättning och borttagande av baspar

När DNA kopieras händer ibland att en eller ett par baser tappas bort eller läggs till. Man kallar sådana händelser för deletioner respektive insertioner. I människans genom finner man att det under de senaste 70 miljonerna år (sedan vi skildes från musen) för vart tusental baspar människa har skett åtta sådana tillägg eller borttaganden, vilka sammantaget lagt till sex och tagit bort 18 baspar.



Figur 4.2. Steroidreceptorer förenar i samma protein tre separata biokemiska aktiviteter, som utförs av varsin domän hos proteinet.. Bild: Henrik Brändén, Vetenskapsrådet.

Frekvensen av sådana mutationer skiljer sig dock kraftigt mellan olika arter och utvecklingslinjer. Under samma tid har det i utvecklingen mot mus skett tre gånger så mycket insertioner och deletioner.

Om ett eller några baspar däremot sätts in eller tas bort inne i en proteinbeskrivande gen är sannolikheten mycket stor att genen slås ut: så länge det insatta eller borttagna antalet baser inte är delbart med tre kommer nämligen läsramen för proteinet att förskjutas, vilket vanligen leder till att en stopptriplett uppkommer mitt inne i genen. De flesta deletioner och insertioner inne i gener rensas därför bort av det naturliga urvalet.

Utbyte av DNA-bokstäver

Ännu vanligare än insättningar och borttaganden av baspar är att en DNA-bokstav helt enkelt byts ut mot en annan, något som kallas för basparsutbyten eller substitutioner. Dessa beror i grund och botten på att DNA-polymeras inte är perfekt och därför ibland fogar in en felaktig DNA-bokstav. Detta kan ske i två sammanhang. Dels när DNA-molekylen ska kopieras, dels när en skada på DNA-molekylen ska repareras. (DNA-molekylen skadas ständigt av spontana kemiska processer, av UV-strålning och av olika aggressiva ämnen i kroppen. Olika reparationsenzymer klipper därför kontinuerligt bort skadade delar av DNA-strängar och överläter åt DNA-polymeras att fylla igen de hål som då uppkommer.) Misstag vid kopiering och misstag vid reparation bidrar i ungefär lika utsträckning till skillnader i DNA-sekvenser mellan individer och arter.

Hastigheten för basparsutbyten varierar kraftigt mellan olika utvecklingslinjer och perioder, så att musen exempelvis samlat på sig nästan lika många basparsutbyten sedan hennes förfäder för femton miljoner år sedan skiljde sig från råttans som människan samlat på sig under de runt 70 miljoner år som gått sedan hennes förfäder separerades från gnagarnas. Frekvensen av denna typ av mutationer under de senaste 70 miljonerna års utveckling mot människa har varit 16%. Under samma tid har den i utvecklingslinjen mot mus varit 31%.

De basparsutbyten som träffar introner och områden mellan gener tycks i de flesta fall tolereras av det naturliga urvalet. Av de basparsutbyten som ändrar en aminosyra i en proteinkodande gen rensas däremot runt 80% bort av det naturliga urvalet inom loppet av 10 miljoner år.

5. Skillnader mellan människors genom

De senaste årtiondets genomforskning har dramatiskt ökat våra kunskaper om vilka skillnader som finns mellan olika människor på DNA-nivå. Om vi bortser från det som åstadkoms av skillnaden mellan en X och en Y-kromosom har tre huvudsakliga typer av skillnader konstaterats:

Skillnader i enstaka baspar (SNPs)

Mellan olika människors genom finns mängder av skillnader i enstaka baspar. Så att en person på ett visst ställe på sin ena kromosom sex har ett A medan han på sin andra har ett C. Medan en annan person kanske på detta ställe har ett A på bägge sina kromosomer. Molekylärbioologer kallar sådana ställen för SNP (single nucleotide polymorphism). Eftersom skillnader i enstaka baspar inte bara finns mellan individer utan också mellan två kromosomer i ett kromosompar mäter man deras frekvens genom att jämföra två enkla kromosomuppsättningar med varandra. Man finner då att ungefär en DNA-bokstav av 1300 skiljer sig mellan två mänskliga kromosomuppsättningar.

De allra flesta sådana ställen åstadkommer inga märkbara skillnader mellan individerna, eftersom de ligger inne i introner eller mellan gener. I några fall kan skillnaden dock ligga inne i en gen, och därmed åstadkomma två olika genvarianter som ger upphov till olika proteiner med skilda funktioner. (Kanske ett som gör blått respektive ett som gör brunt ögonpigment). I andra fall kan skillnaden ligga inne i en styrsekvens, som avgör hur mycket en gen ska användas. Sådana skillnader kan påverka allt från hudfärg (genom att styra hur mycket man bildar av proteiner som tillverkar pigment) till skillnader i reglering av humör (exempelvis genom att påverka hur mycket man bildar av enzymer som tillverkar eller bryter ner olika signalämnen i hjärnan).

Duplikationer

När man jämför arvsanlagen mellan två människor hittar man tiotals regioner om flera tusen baspar vardera som är duplicerade på någon kromosom hos den ena hos den andra, men inte hos den andra människan. Det är i skrivande stund inte känt exakt hur många och hur stora dessa regioner är, men man vet att ett antal proteinbeskrivande gener finns i de duplikat man redan idag upptäckt. I några fall har man sett att sådana duplikationer av områden med en eller flera gener kan åstadkomma allvarliga medfödda syndrom och sjukdomar. I övrigt vet man inte i vilken utsträckning sådana duplikationer

bidrar till fysiska skillnader mellan olika individer. De flesta gener antas nämligen ha en feedback-reglering, där användningen av genen inte bara anpassas efter behovet av dess protein utan även av mängden man redan har av proteinet. Därigenom blir mängden man tillverkar av ett protein inte beroende av antalet kopior av dess gen.

Olika längd på enkla repeats (mikrosatelliter)

Utspridda på vårt genom finns dessutom massor av ställen (ett vart tvåtusende baspar) där korta DNA-sekvenser (1-13 baspar) upprepas gång på gång efter varandra. Eftersom DNA-polymeras har lätt att "slira" på sådana ställen, och av misstag föra in för få eller för många av dessa repeterade sekvenser, har det uppstått en kraftig variation mellan olika individer i längden på dessa ställen. På ett sådant ställe skulle exempelvis en person på sin ena kromosom kunna ha fem och på den andra 28 repeterade enheter, medan motsvarande siffror för en annan person skulle kunna vara 14 och 19. Dessa ställen kallas av molekylärbioologer omväxlande för mikrosatelliter och hypervariabla tandemrepeats.

De allra flesta mikrosatelliter ligger mellan gener eller i introner, och skillnader i längden på dessa spelar därför ingen roll för våra egenskaper. (Undantaget är några ärftliga sjukdomar där en tre baser lång repeterad sekvens ligger inne i en gen. Därmed får proteinet en rad identiska aminosyror efter varandra, och exempelvis Huntingtons sjukdom uppkommer om raden av sådana upprepade aminosyror blir för lång.) Eftersom längden på mikrosatelliterna varierar mycket kraftigt mellan olika individer utnyttjas analyser av dessa som genetiska fingeravtryck i rättsväsendet.

Selektion hos människan

Vi ser alltså att olika människor har olika regioner av sitt DNA duplicerat och vi ser också att små mutationer gör att den exakta ordningsföljden av DNA-bokstäver kan skilja sig mellan olika individer. Därav kan man sluta sig till att det inom människoarten finns och ständigt har funnits en genetisk variation, som det naturliga urvalet kan arbeta med. Forskarna har då frågat sig i vilka tidsperspektiv det naturliga urvalet verkat. Och kommit fram till att tio- eller hundratals nya genvarianter uppkommit i olika grupper av människor under de senaste tiotusen åren, som visat sig ha så starkt överlevnadsvärde att de blivit

utsatta för positiv selektion. Inte nödvändigtvis så att den nya genvarianten helt trängt undan den gamla, men i vart fall så att frekvensen av den nya genvarianten blivit betydligt högre än slumpen skulle kunna förklara.

- Många av dessa gener beskriver proteiner som är inblandade i att känna igen lukter och smaker. Vilket inte är så märkligt, med tanke på att lukt och smak både hjälper oss att välja ut lämpliga födoämnen och att upptäcka och undgå faror. Och att både tänkbara födoämnen och faror varierar med tid och plats. Så att förmåga att känna igen nya lukter och smaker rimligen haft stor betydelse för att anpassa människan till olika miljöförändringar, såväl sådana som orsakas av yttre omständigheter som klimatväxlingar eller byte av bostadsort, och sådana som orsakas av ändrat mänskligt beteende, som övergång till jordbruk. Något som skett just under de senaste tiotusen åren.
- Ett betydande antal av generna beskriver proteiner inblandade i produktion av könsceller och i befruktning.
- Vidare beskriver många av generna proteiner som är inblandade i ämnesomsättningen. Många gener deltar i nedbrytningen av kolhydrater och fetter. Exempelvis har europeer selekterats för en gen som hjälper oss hantera mjölksocker (laktos). Andra gener i syntesen av vitaminer. I Asien har människor selekterats för en speciell variant av genen för alkoholdehydrogenas. Forskarna utgår från att selektion av sådana gener är en del av anpassningen till ett liv som jordbrukare, där olika grödor i olika regioner lett till olika selektiva tryck för olika typer av genetiska förändringar.
- Även gener som påverkar utseende tycks ha varit utsatta för selektion hos människan. Den mest uppmärksammade skillnaden mellan människor från olika delar av världen är hudfärg, och den bestäms av mängden i huden av färgämnet melanin. Denna mängd har det naturliga urvalet anpassat till olika nivåer hos olika folkgrupper i en avvägning mellan två krav: För det första är solens UV-strålar farliga för cellerna, och melanin skyddar mot UV. För det andra behöver jordbrukande men inte jagande-samlade människor trots detta en viss mängd UV-ljus, för att omvandla ett förstadium av vitamin A (som

finns i många grödor) till fungerande vitamin. Så ju längre från ekvatorn jordbrukande människor lever, desto större selektivt tryck för låg melaninproduktion. Dessa förändringar kan ha gått mycket snabbt. När delar av det judiska folket tvingades från medelhavsområdet norrut tog det betydligt mindre än 2000 år innan deras hudfärg blivit likadan som nordeuropeernas. Medan de trosfränder som stannade kvar i medelhavsområdet behöll sin mörkare hy. Och de som flydde söder om Sahara utvecklade betydligt mörkare hy. Tester av andra DNA-varianter visar att detta inte varit resultatet av att judarna blandats upp av omkringliggande folk, utan av naturligt urval.

Även andra skillnader i utseende mellan olika folkgrupper beror på små genetiska anpassningar till olika klimat. Kompakt kroppsbyggnad ger stor kroppsvolym i förhållande till kroppsyta, och gynnas i klimat där värmeisolering är viktig. Spenslig kroppsbyggnad (stor yta i förhållande till volym) gynnas i klimat där avkylning är viktigt. På liknande sätt kan klimatet gynna utvecklingen av olika form på näsa och hår. Många sådana skillnader kan mycket väl ha uppkommit mycket längre tillbaka i tiden än de senaste 10 000 åren, men även bland de gener som varit utsatta för selektion de senaste 10 000 åren hittar man sådana som styr tillväxten hos olika ben och därmed kontrollerar formen hos olika delar av kroppen.

Det naturliga urvalet har alltså gång på gång sedan jordbrukets uppkomst slipat och anpassat vår genuppsättning. Och det är egentligen inte överraskande, ty jordbruket har inneburit dramatiska förändringar av människans miljö – både vad gäller de faror vi möter och den föda vi äter. När forskare nu upptäcker att människor har olika varianter av gener som påverkar risken för olika sjukdomar som diabetes och hjärtinfarkt, och upptäcker att högriskgenerna är speciellt farliga i kombination med en ”modern” kosthållning, är det i många fall en ögonblicksbild av en pågående utveckling, där gener anpassade efter en icke-jordbrukande miljö byts ut mot gener bättre anpassade för den moderna människans livsföring. Ofta visar det sig när man undersöker sådana ”högriskgener” att högriskvarianten av genen är identisk med den gen som finns hos schimpansen, medan ”lågriskvarianten” är en mänsklig innovation.

6. Skillnader mellan människans och schimpansens genom

Schimpansen är människans närmaste släkting, och forskarna uppskattar att de två arterna separerades för ungefär sex eller sju miljoner år sedan. Analyser av skillnaderna mellan schimpansens och människans genom kan därför kasta ljus över vad för typer av genetiska förändringar som medverkar till separationen av arter, och hur stora skillnader som hinner utvecklas inom loppet av en handfull årmiljoner.

Storskaliga kromosomskillnader

Sedan separationen mellan schimpans och människa har i utvecklingslinjen mot människa två kromosomer slagits ihop till en. Den kromosom som hos människan kallas nr 2 motsvarar därför hos schimpansen två kromosomer, numera kallade 2A och 2B. Därtill har det under utvecklingen mot människa eller schimpans på nio ställen skett så kallade inversioner, där en betydande del av kromosomen vänts bakfram.

Dupliceringar

Under utvecklingen från den gemensamma förfadern till människa har vidare duplikationer skapat dubbla kopior i människans genom av över 500 regioner om i genomsnitt 55 tusen baspar. Åtminstone 88 proteingener finns i dessa regioner. Hos schimpansen har på motsvarande sätt dupliceringar skett av 200 regioner om ungefär samma storlek. För några av de gener som ligger i sådana duplicerade regionerna kan man se att mängden som bildas av proteinet varierar med antalet kopior av genen.

Transposoner

Transposoner har åstadkommit många skillnader mellan schimpansens och människans genom. En grupp av transposoner som kallas LINEs har stoppat in ca 2000 ytterligare kopior av delar av sig själv på olika platser i genomen under utvecklingsvägen mot schimpans, och lika många kopior på andra platser i genomet under utvecklingsvägen mot människa. En annan grupp av transposoner som kallas SINEs har åstadkommit 7000 liknande instoppningar av delar av sig själv under utvecklingen mot människa, men bara drygt 2000 under utvecklingen mot schimpans. En tredje grupp av transposoner som kallas LTR har i det närmaste avsmnat hos människan, medan schimpansen tycks ha tagit emot två nya sådana transposoner, som skapat sammanlagt 200 kopior i schimpansens DNA. Dessa nya transposonkopior utgör dock bara en

bråkdel av människans och schimpansens totala antal transposonkopior. De allra flesta transposonkopior i vårt genom finns därför på samma plats hos schimpansen.

Skillnader i enstaka baspar

Om vi bortser från de skillnader som åstadkommes av dupliceringar och instoppningar av hela eller delar av transposoner skiljer sig DNA-sekvensen mellan en kromosomuppsättning hos schimpansen och en kromosomuppsättning hos människan på 1,23% av basparen. Skillnaden i DNA-sekvens mellan en människa och en schimpans är således ungefär 10 gånger så stor som skillnaderna mellan två människor. Och runt fem gånger så stor som skillnaden mellan två schimpanser.

Skillnader på proteinnivå

Vad åstadkommer då detta för skillnader mellan människans och schimpansens uppsättning av proteiner?

För det första kan man se att över 30 av de proteingener man hittat hos människan tycks saknas hos schimpansen. Bland dessa märks bland annat gener för ett antal proteiner, som är inblandade i kroppens försvar mot olika sjukdomsalstrare. Däremot inga som direkt för tankarna till de mentala förmågor, som vi anser vara de viktigaste skillnaderna mellan oss och schimpansen. Det är mycket troligt att man med tiden kommer att hitta en motsvarande uppsättning proteiner, som finns hos schimpansen men saknas hos människan.

För det andra kan man se att det hos 70 % av de proteiner vi har gemensamt med schimpansen finns små skillnader i aminosyresekvensen. Oftast är det då en eller två aminosyror som skiljer. De flesta gener tycks ha varit utsatta för negativ selektion, och har färre mutationer som orsakar aminosyreutbyten än den generella mutationsfrekvensen skulle åstadkommit. Speciellt stark tycks den negativa selektionen ha varit i gener för proteiner som är inblandade i signalkickande inne i cellerna, metabolism, neurogenes och förmedling av signaler vid synapser. Några grupper av gener visar dock tecken på att i stället ha varit utsatta för positiv selektion (vilket man ser på att de uppvisar fler aminosyresubstitutioner än den generella mutationsfrekvensen skulle ge skäl att förmoda). Dit hör en rad gener inblandade i försvar mot olika

sjukdomsalstrande mikroorganismer, i reproduktion och i uppfångande av sinnesförmågelser (främst lukter).

För det tredje kan man utgå från att en del av skillnaderna i DNA-sekvensen befinner sig i de styrsekvenser, som påverkar var, när och hur mycket olika gener avläses. Denna förmodan styrks av storskaliga undersökningar som visar att det för många gener finns betydande skillnader mellan arterna i mängden mRNA som bildas från genen. Skillnaderna är dessutom olika stora i olika vävnader: I hjärnan hittade forskarna sådana skillnader för ungefär 8% av generna, i testiklarna för hela 32% av generna. Andra vävnader befinner sig däremellan. Hjärnan är alltså det organ i kroppen där skillnaderna i genuttryck mellan människa och schimpans är som minst!

Vi kommer alltså till slutsatsen att skillnaderna mellan människa och schimpans består i att några enstaka finns hos den ena men inte den andra arten. Och att skillnaderna i övrigt ligger hos enstaka aminosyror hos våra proteiner och i de reglersekvenser, som styr hur mycket vi bildar av olika proteiner. Det är dock fullt möjligt att sådana till synes små genetiska skillnader kan åstadkomma dramatiska skillnader i funktion. Till exempel vet vi från andra arter att enstaka baspars skillnad i styrsekvenser för proteiner som under foster- och embryoutveckling talar om för olika grupper av celler om de ska dela sig eller ej kan få dramatiska effekter på individens kroppsbyggnad och utseende.

Vad gör oss mänskliga?

Kraftigast skillnader mellan människa och schimpans hittar man alltså i gener som deltar i luktsinnet, fortplantningen och försvaret mot främmande parasiter. Detta är inte konstigt: Både människan och schimpansen har levt och lever i en ständigt föränderlig miljö där man möter nya parasiter, rovdjur och andra hot. Vilket gör att nya genvarianter bevaras av det naturliga urvalet ifall de ger sin bärare bättre möjligheter att hantera dessa hot. Eller, förstås, om de gör sin bärare bättre på att föröka sig.

Det är dock varken vår förmåga att hantera en unik uppsättning av mikroorganismer, våra säregna sexualvanor eller vår förmåga att uppfatta något annorlunda lukter än schimpansen som vi betraktar som de drag som gör oss till människor. Man brukar istället peka på vår ovanligt stora hjärna, vår förmåga till abstrakt språk och vår upprättgående gång. Vilka genetiska förändringar som givit oss dessa förmågor är fortfarande en obesvarad fråga, men några ledtrådar är kända:

- Det finns två gener (kallade ASPM och MCPH1) som tycks spela en viktig roll för att hjälpa de celler som bildar hjärnans nervceller att dela sig. Människor där en av generna inte fungerar drabbas av förkrymt hjärna (mikroencefali) med ungefär hälften så stor hjärna som normalt. Mellan schimpans och människa finns små skillnader i genen för detta protein, och tecken tyder på att det naturliga urvalet kan ha selekterat för dessa skillnader.
- Om genen FOXP2 är muterad hos människor får man en allvarlig störning av både artikulation och språkförmåga. Genen skiljer sig något mellan människa och schimpans. Även här tyder analyser av mängden aminosyreutbyten att genen kan ha varit utsatt för positiv selektion. Dessutom antyder andra analyser att den genen (eller någon annan i dess närhet) för några hundra tusen år sedan råkade ut för en förändring som gjorde att den förändrade genen gav en så stark överlevnadsfördel, att den snabbt trängde undan den tidigare varianten av genen.
- Genen AHI1 behövs för att axoner ska finna rätt väg från hjärnans cortex till ryggmärgen. Om genen är muterad hos människor uppkommer en nervsjukdom som både påverkar kroppshållning och ger mental retardation, autistiska symptom och antisocialt beteende. Även hos denna gen finns små skillnader mellan människor och primater, och data antyder även här att positiv selektion kan ha förekommit. I så fall skulle det kunna ha varit en genetisk förändring som samtidigt bidragit till upprättgående gång, språklig förmåga och ett specifikt socialt beteende.

7. Skillnader mellan människans och gnagares genom

De första däggdjursgenom som sekvensbestämdes efter människans var musens och råttans. Dessa gnagare anses ha separerats från människans utvecklingslinje för mellan 65 och 75 miljoner år sedan, medan den sista gemensamma förfadern till både råtta och mus uppskattas ha levt för mellan 12 och 24 miljoner år sedan.

Storskaliga skillnader

Genomets storlek är ungefär detsamma hos alla däggdjur. Både hos människa, mus och råtta omfattar eukromatinet mellan 2,6 och 2,9 miljarder baspar. Antalet kromosomer och deras bandmönster skiljer sig dock dramatiskt från varandra. När man analyserar vilka gener som finns på kromosomerna och i vilken ordning de ligger kan man dock i människans, musens och råttans kromosomer urskilja drygt 300 mindre delar (kallade syntena block), där samma gener ligger i samma ordning efter varandra inom ett block, men där blocken är hopfogade på olika sätt till olika antal kromosomer i de olika arterna. Detta kan förklaras med att genomen under utvecklingen utsatts för en rad olika rearrangeringar, där segment av kromosomer bytt plats eller vänts bakfram. Under utvecklingen mot människa, mus och råtta har det behövt ske minst 350 sådana rearrangeringar för att ge upphov till dagens kromosomarrangemang hos de tre arterna. Flertalet av dessa rearrangemang har ägt rum i utvecklingslinjen mot dagens gnagare.

Dupliceringar

Alla de duplikationer som diskuterades i diskussionen om människors och schimpansers genom har uppkommit efter det att primaternas och gnagarnas förfäder separerades. På samma sätt som duplikationer skett under utvecklingen mot primater har de naturligtvis också skett under utvecklingen mot gnagare, men det tycks som om en något lägre andel av gnagarnas än primaternas genom utgörs av duplikationer. Medan ungefär 5%

Transposonklass	Human	Mus	Råtta
LINEs	13%	20%	23%
SINEs	20%	8%	7%
LTR-element	8%	10%	9%
DNA-transp.	3%	1%	1%
Totalt	44%	39%	40%

Tabell 7.1 Andel av genom som utgörs av olika klasser av transposoner hos människa, mus och råtta.

av människans genom består av duplikationer är siffran 3% för råtta och 1-2% för mus.

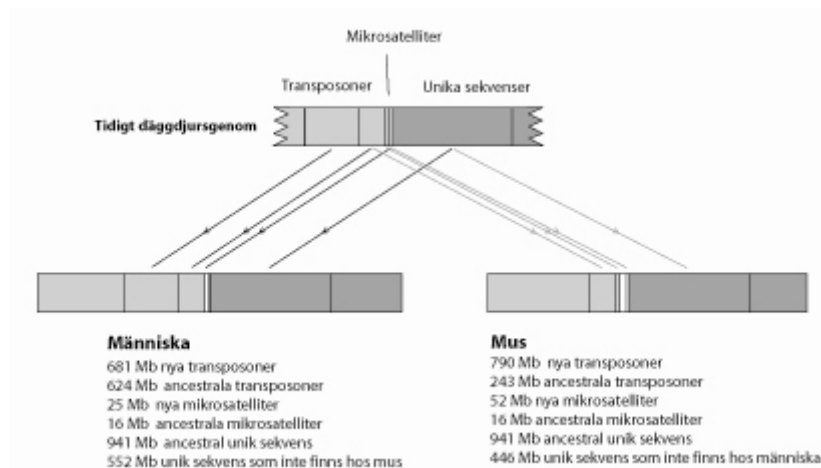
Transposoner

Hos både mus och råtta är transposonernas andel av genom något lägre än hos människan. Detta trots att transposoner hoppat betydligt mer under utvecklingen mot gnagare än under utvecklingen mot människa. Men eftersom även basparsutbytena varit fler hos gnagarna har helt enkelt en massa transposonsekvenser hos mus och råtta muterats till oigenkännlighet, medan de fortfarande går att identifiera hos människan. Detta har lett till att bara en åttondel av musens igenkännbara transposoner men hälften av människans funnits redan innan människans och gnagarnas förfäder särades från varandra.

Eftersom hälften av människans transposoner, och den överväldigande majoriteten av gnagarnas, hoppat efter splittringen mellan arterna är det bara en liten andel av transposonerna som sitter på samma ställe i de två arterna. Dessutom har det funnits betydande skillnader i vilka typer av transposoner som varit mest aktiva i de två utvecklingslinjerna. Vilket lett till att den relativa andelen av olika transposonklasser skiljer sig mellan människa och gnagare (tabell 4.1).

Däggdjurs genomkärna

Genom att jämföra människans, musens och råttans genom kan man skaffa sig en uppfattning om hur genom kan ha sett ut hos de tidiga däggdjuren (figur 7.1). Man finner då att huvuddelen av de sekvenser som ligger utanför transposoner är gemensamma för både människa och mus, och i denna gemensamma kärna på en knapp miljard baspar finns i stort sett alla protein-gener hos såväl människa som mus. Vidare kan man som redan antytts hos människan fortfarande urskilja en betydligt större andel av de transposoner som fanns hos det tidiga däggdjuret än hos musen. Men hos både människa och mus har en stor mängd ny sekvens uppkommit genom att transposoner hoppat. En betydande andel av både musens och människans genom utgörs också av unika sekvenser som inte förekommer hos den andra arten. Delar av detta kan vara regioner av det ursprungliga däggdjursgenomet, som deleterats i den ena men inte i den andra arten. Andra delar kan vara transposoner, som muterats så mycket att



Figur 7.1. Människans och musens genom och de delar dessa måste ha gemensamt med sin gemensamma förfader (kallat tidigt däggdjursgenom).

de inte längre går att känna igen. Hur mycket transposoner och unik sekvens som fanns i det ursprungliga däggdjursgenomet men som försvunnit under utvecklingen mot både gnagare och människor är okänt.

Småskaliga DNA-skilnader

I de områden av musens och människans genom som är gemensamma men som inte varit utsatt för selektion finner man för varje tusental baspar i genomsnitt 35 insertioner och deletioner, som omfattar sammanlagt 80 baspar. Huvuddelen av dessa har skett i utvecklingslinjen mot mus. Detta har lett till att det i musens genom finns fyra eller fem gånger så många mikrosatelliter som i människans genom.

Bortsett från dessa instoppningar och borttaganden av baspar finner man att själva DNA-sekvensen i dessa icke-selektade regioner av genomet skiljer sig på ca 40% av DNA-bokstäverna. Eftersom somliga kvävebaser förändrats flera gånger under denna utvecklingslinje avspeglar detta en något högre frekvens av mutationer under utvecklingen mellan människa och mus. I de DNA-sekvenser som beskriver proteiner är frekvensen av basparutbyten betydligt lägre eftersom det naturliga urvalet rensat bort de flesta sådana som lett till förändrade aminosyror.

Skilnader på proteinnivå

Mus och råttor har i stort sett lika många gener som människa. Runt 90% av generna hos råttan har en direkt och entydig motsvarighet hos människan. Nästan alla resterande gener hör till olika genkluster, där dupliceringar både före och efter splittringen mellan människa och mus lett till att

proteiner, som påminner mycket om varandra, men både människan och musen har flera olika som inte går att entydigt para ihop mellan arterna. Det är framför allt fyra grupper av gener som på detta sätt duplicerats och utvecklats flitigt både före och efter splittringen mellan människa och mus.

Nämnligen gener som beskriver:

- Luktreceptorer. Proteiner som sitter i luktepitelet i näsan och kan fånga upp olika ämnen.
- Avgiftningenszymer. Närmare bestämt proteiner kallade p450 som finns i levern och bryter ner olika giftiga ämnen.
- Proteiner som är inblandade i fortplantningen.
- Proteiner som deltar i vårt försvar mot bakterier, virus och andra parasiter.

Detta avspeglar naturligtvis att urvalet gynnar dem som blir bättre på att skaffa barn och klarar av att anpassa sig till nya miljöer. Nya försvarsproteiner hjälper naturligtvis till att hantera nya parasiter. Förmåga att uppfatta nya lukter hjälper som redan nämnts att undvika både rovdjur, giftiga födoämnen och andra faror i nya miljöer. Nya avgiftningenszymer hjälper en att hantera nya födoämnen som annars skulle vara giftiga.

Studerar man sedan gener för proteiner som är varandras direkta motsvarigheter mellan människa och råttor eller människa och mus ser man att likheten i aminosyresekvens är i genomsnitt 88%. Om man analyserar olika grupper av proteiner med olika funktion ser man dock tydliga variationer av denna siffra. Hos proteiner inblandade i exempelvis luktsinne, försvar och reproduktion är skillnaderna större, hos proteiner inblandade i ämnesomsättning och cellers inre signalkickande är skillnaderna mindre.

8. Skillnader mellan däggdjur och andra arter

Likheterna är alltså mycket stora mellan människa och andra däggdjur när man undersöker både hur genomet ser ut, och vad det innehåller. Ju längre bort man kommer från människan i fylogenin, desto större blir dock skillnaderna.

Genomets storlek

Genomets fysiska storlek varierar dramatiskt mellan olika arter. Visserligen finns en tendens att genomet blir mindre ju ”enklare” vi människor uppfattar arten vara, men sambandet är ingalunda entydigt: Bönor har med sina 30 miljarder baspar i ”enkel” kromosomuppsättning tio gånger mer DNA per cell än en människa. Risplantans celler har däremot bara en sjättedel av människocellernas DNA-mängd. De små zebrafiskarna har fyra gånger så mycket DNA som de fugu-fiskar, vars giftblåsor gör japanernas ätandet av dessa delikatesser till en form av rysk roulette. Den encelliga protozoen *Amoeba dubia* har 200 gånger större genom än människan.

Antal gener och gentäthet

Medan mängden DNA varierar kraftigt mellan olika arter är variationen i antalet proteingener betydligt mindre. Flest tycks växter ha, man tror att risplantan har över 30 000 proteingener och hos backtrav (*Arabidopsis thaliana*) har man hittat över 27 000. Alla däggdjur har ungefär 22 000 gener som beskriver proteiner. När man undersökt genomet hos kyckling och ett par fiskar finner man ungefär samma siffra. När man däremot kommer till de ryggradslösa djuren sjunker antalet proteingener något: Hos rundmasken *C. elegans* finns 19 000, hos bananflugan 14 000. Malariaparasiten har drygt 5000 proteingener. Malaria-parasiten

tycks ha drygt 5000 proteingener, *E. coli*-bakterien 4500 och den enklaste bakterie man känner till, *Mycobacterium genitalium*, knappt 500.

Gentätheten varierar således kraftigt mellan olika arter. Rundmasken *C. elegans* har ungefär lika många gener som människan på ett genom som är mindre än en tjugondel så stort. I människans genom finns 136 000 baspar för varje proteingen. Hos jästsvampen är motsvarande siffra 2200 och hos *E. coli*-bakterien ungefär 1100 baspar. Det kan jämföras med att genomsnittliga proteiner består av mellan 300 och 500 aminosyror, och att beskrivningarna av dem således kräver 900 till 1500 baspar. Hos människan upptas ungefär 1,5% av genomet av beskrivningar av aminosyrekedjor. Hos bananflugan är siffran 10% och hos många bakterier långt över 90%.

Transposoner och introner

Dessa skillnader i gentäthet har ett samband dels med mängden döda transposoner i genomet (ju tätare generna ligger, desto mindre plats finns i genomet för ett stort antal transposoner), dels antalet introner och deras längd. (Se tabell 8.1.) Hos bakterierna, där generna ligger mycket tätt, finns inga introner. I jästsvampen, där generna ligger något glesare, har fem procent av generna introner. Hos malariaparasiten har över hälften av generna introner, och det genomsnittliga antalet introner är 2,5 per gen. Ryggradslösa djur har dubbelt så många introner, och ryggradsdjuret har alla ungefär åtta introner per gen. Längden på intronerna varierar ännu kraftigare än deras antal, och samvarierar tydligt med gentätheten. Hos malariaparasiten är den genomsnittliga intronen

Organism	Genomstorlek (miljoner baspar)	Antal proteingener	Antal baspar /gen	Andel transposon	Andel gener med introner	Antal introner /gen	Intronernas längd (baspar)
Människa	2 900	22 000	132 000	0,44	> 0,9	8	3 400
Råtta	2 750	22 000	125 000	0,40	> 0,9	8	3 200
Mus	2 600	22 000	118 000	0,39	> 0,9	8	3 100
Kyckling	1 200	22 000	55 000	0,09	> 0,9	8	1 500
Fugu	400	22 000	18 000	0,03	< 0,9	8	700
Bananfluga	137	14 000	9 800	0,03	Flertalet	7	490
Rundmask	97	19 000	5 100	0,07	Flertalet	6	270
Backtrav	125	27 000	4 600	0,14	0,79	5	170
Malariaparasit	23	5 300	4 300	-	0,54	2,5	180
Jäst	12,5	5 800	2 200	-	0,05	-	-
<i>E. coli</i>	4,7	4 300	1 100	-	0	-	-
<i>Mycobacterium genitalium</i>	0,58	480	1 200	-	0	-	-

Tabell 8.1. Genomstorlek, geninnehåll, transposoner och introner i olika arter

kortare än 200 baspar, hos däggdjuren längre än 3000 baspar.

Sekvenslikheter

I regioner som, inte varit utsatta för något selektivt tryck att bevaras går det inte att se någon som helst sekvenslikhet mellan däggdjur och andra arter. Redan mellan människa och kyckling har det skett i genomsnitt två substitutioner per baspar. Däremot kan man iaktta mycket avlägsna släktskap om man jämför sekvenser inne i gener, och hos de proteiner dessa beskriver. Hos proteiner som är varandras motsvarigheter är exempelvis 75% av aminosyrorerna identiska mellan kyckling och människa. För de starkast konserverade proteinerna, exempelvis enzymer inblandade i de mest grundläggande delarna av ämnesomsättningen, kan man till och med se tydliga likheter mellan de tre olika "superkungarikena" eukaryoter, arkebakterier och eubakterier (se tabell 8.3).

Proteinuppsättningar

Även uppsättningen proteiner skiljer sig allt mer mellan olika arter ju avlägsnare de är släkt med varandra. Medan vi sett att 90% av musens gener har en entydig motsvarighet hos människan är siffran för hönan 60%. Av de återstående generna har dock mer än hälften en motsvarighet som inte är entydig. Vilket innebär att dessa gener har skapats genom dupliceringar efter det människan och hönan skiljdes åt.

15% av hönans gener saknar dock motsvarighet hos människan. Och ungefär lika stor andel av människans gener saknar motsvarighet hos kycklingen. I de flesta fall är det då frågan om gener som duplicerats och börjat utvecklas åt olika håll långt innan människans och hönans förfäder skilde sig åt, men som sedan försvunnit i den ena men inte i den andra arten. I en del fall är det till och med frågan om att hela proteinfamiljer försvunnit i en av arterna. Exempelvis finns hos

Organism	Trios fosfat isomerer	Pyruvat kinas	Fumaras	Alkohol dehydrogenas
Mus	0,95	0,97	0,96	0,89
Kyckling	0,89	0,86	0,89	0,69
Fluga	0,65	0,65	0,76	0,52
Mask	0,62	0,59	0,76	0,51
Jäst	0,53	0,51	0,67	0,46
Malaria-Parasit	0,43	0,44	-	-
E. coli (eubakterie)	0,45	0,47	0,60	0,46
Halobakterie (arkae-bakterie)	0,15	0,38	0,50	0,26

Tabell 8.3. Likhet i aminosyressekvens mellan människa och andra arter hos några viktiga enzymer.

kycklingen en speciell familj av keratinproteiner, som helt saknas hos människan, som behövs i näbb, fjädrar, klor och äggskal. Däggdjuren har å sin sida några familjer med bland annat mjölkproteiner, som förlorats hos fåglarna.

Jämför man däggdjur med ryggradslösa djur blir skillnaderna ännu större. Hos bananflugan saknar 40% av generna en motsvarighet hos människan, och hos nematoder (rundmaskar) mellan 50 och 60%. Här är det inte bara frågan om att gener i den gemensamma förfadern råkat ut för olika öden (duplicering eller eliminering) under utvecklingen. Utan också om att gener och delar av gener förts ihop till nya arrangemang av domäner under utvecklingen av de olika djurgrupperna. Särskilt intensivt tycks denna verksamhet ha varit i utvecklingslinjen mot ryggradsdjur, som har ungefär dubbelt så många olika domänarrangemang som insekter eller maskar. Under denna utveckling har även omskyfflingar av enstaka exoner skapat en handfull nya domäntyper, som förekommer hos ryggradsdjur men inte hos ryggradslösa djur.

Andra eukaryota varelser skiljer sig naturligtvis ännu mer från människan. Visserligen har bortåt hälften av jästsvampens proteingener en motsvarighet hos människan, men det rör sig i absoluta tal om mindre än 3000 gener. Även under utvecklingen från de tidigaste eukaryoterna till de första djuren måste ett stort antal nya domänarrangemang ha skapats, ty hos både insekter och maskar finns dubbelt så många olika sätt som hos jäst att arrangera domäner till proteiner.

Ett annat sätt att beskriva skillnaderna i proteinuppsättningar mellan olika grupper av arter är att utgå från människans genom och fråga sig hur många av dess gener som har tydliga motsvarigheter hos andra grupper av levande varelser (tabell 4.4). Man finner då att nästan en fjärdedel av våra gener även förekommer även hos bakterier. Hälften av dem har hittats hos någon svamp eller protozo. Tre fjärdedelar av dem finns i ett eller flera ryggradslösa djur. Endast en procent av våra gener är specifika för primater.

Andel av människans gener	Motsvarighet finns hos
23%	Bakterier
50%	Svampar och protozoer
78%	Ryggradslösa djur
85%	Fiskar och fåglar och reptiler
99%	Däggdjur utom primater

Tabell 8.4. Andel av människans gener som har direkt motsvarighet i olika livsformer.

